DOCUMENT RESUME

ED 202 214 FL 012 270

AUTHOR Hawkins, Sarah; Tatham, M. A. A.
TITLE On the Regulation of Consonant Duration in the Speech Perception and Production of Adults and 5 to 7 Year Old Children.
INSTITUTION Essex Univ., Colchester (England). Dept. of Language and Linguistics.
PUB DATE Apr 80
NOTE 71p.; In its Occasional Papers, 23, Phonetics and Phonology, p1-70, Apr 1980.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Adults; *Articulation (Speech); *Auditory Perception; Cognitive Processes; *Comparative Analysis; *Consonants; Psychomotor Skills; Sound Spectrographs; Stimuli; Young Children
IDENTIFIERS *Consonant Cluster Duration

ABSTRACT

The perception of duration in consonant segments is compared with their produced durations, and adults' and children's perfcrmance in this respect are compared. The temporal organization imposed by adults on initial consonant clusters was often not imposed by children from four to eight years of age, even though the cluster was produced "correctly." An experiment was designed to help determine the perceptual importance of the differences between adults' and children's productions, and to indicate some possible causes of the differences. The research questions were: (1) What are the just noticeable differences (JNDs) for durations of consonants in clusters? (2) Are the characteristic durations of clustered consonants in the speech of adults also preferred by them in their perception of speech? (3) Do the immature patterns in children's production of consonant clusters stem from muscular constraints or from an immaturity of the central timing regulator? Tables include an analysis of (1) variance for the variances of stimuli, (2) mean duration and standard deviations in perception and production of /s/ and /t/ by children and adults, and (3) Mann-Whitney U tests on adult and child voice stimuli. Spectrograms of the synthetic speech of the experiments are presented. (JK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Univ. of Essex
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED202214

On the Regulation of Consonant Duration in the Speech Perception and Production of Adults and 5 to 7 Year Old Children

Sarah Hawkins* and M.A.A. Tatham

Final Report for the Science Research Council grant awarded to Prof. M.A.A. Tatham and Dr. Sarah Hawkins in the Department of Language and Linguistics, University of Essex

*Now at the Dental Research Center, University of North Carolina

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Univ. of Essex, Dept. of Language and Linguistics, Occasional Papers 23, Phonetics and Phonology. Apr. 1980

Abstract

A method of adjustment to internal criterion was used to investigate the perception of consonant duration. Three adults and two children, 5 and 7 years old, adjusted the duration of /s/ and /t/ in synthesised words with prestressed singleton and clustered consonants. The words were presented in two synthesised carrier phrases, in both an adult male and a child's voice. The perceptual data were compared with data for production of /s, t/ in the same contexts by the same subjects, and with values for other subjects' production of the same clusters in different contexts. The results showed that the just noticeable difference for consonant segment duration is of the order of 25 ms., but that the criterion for 'acceptable' durations can shift over time by more than the jnd at any one time. Durations of /s, t/ in perception and production were strongly associated within and between subjects, but there were some adult-child differences of interest. In perception, children appear to lack sensitivity to differences between sentence contexts that adults distinguish between in setting 'acceptable' durations of segments. However, children but not adults required different durations for /s, t/ in some contexts for the child speaker compared with the adult speaker. The differences required in perception for the child speaker did not always mirror the child listener's own durations in production. These data are interpreted with respect to possible processes governing the regulation of segmental timing and rhythm in perception and production.

Acknowledgements

We are most grateful to Professor A.C. Gimson and to Dr. A. Fourcin for making available computing and speech synthesis facilities at University College London, to Dr. M.P. Haggard of the Medical Research Council's Perception Unit at Nottingham University for his valuable advice, and to Dr. I. Witten of the Department of Electrical Engineering Science at the University of Essex; also to Dr. G. Allen of the University of North Carolina for detailed discussions of the data. Special apparatus was constructed by the Electronics Centre at the University of Essex.

INTRODUCTION

The aim of the research reported in this paper was to gather information about the perception of duration in consonant segments compared with their produced durations, and to compare adults' and children's performance in this respect. These aims were formulated as a result of investigations of the speech of 4 to 8 year old children, which showed that the temporal organisation imposed by adults on initial consonant clusters was often not imposed by children of this age, even though the cluster was produced 'correctly', (Hawkins, 1976). The experiment to be reported was intended to help determine the perceptual importance of the differences between adults' and children's productions, and to indicate some possible causes of the differences.

The perceptual importance of the adult-child production differences may best be assessed by finding the just noticeable differences (jnd's) for duration of the segments concerned. The causes of the differences are less open to investigation, but some inferences might be made by comparing adult performance in production and perception of the particular sequences with children's perception and production of the same sequences.

Perception of segmental duration in speech

There are very few studies of the perception of duration for speech segments. Extrapolation of results from discrimination experiments with nonspeech sounds is of only dubious value, since while it is generally agreed that we may expect differential sensitivity to speech and nonspeech stimuli, there is little agreement as to which type of sound will produce smaller difference limens. For example, in discussing absolute difference limens for the durations of nonspeech sounds and their relevance to speech duration discrimination, Lehiste (1970) writes

> "It is quite likely that in a speech situation, where a great amount of external noise is present, the perception would not be as acute." (1970, p.12)"

Rudolph, on the other hand, argues

> "It does not seem likely, however, that a noisy situation would change the difference limens that much. The present results suggest that, at least in a laboratory situation and with synthetic speech, duration discrimination may be at least as accurate as suggested by Fehr et al., and more accurate than one would predict from the results of some older psychophysical studies of duration discrimination in non-speech sounds ...." (1971, p.19).

The experiment by Fehr, Fenske, Fillmore, Langer and Rose (1968) gave absolute difference limens for duration between 1.7 and 5.9 ms. for durations between 10 and 100 ms., whereas LaRiviere (1975) suggested that in normal speech, jnd's would be of the order of 10 to 40 ms.

Rudolph (1971, experiment 1) used synthesised words of from 1 to 4 syllables, presented in isolation. Subjects adjusted the durations of vowels in the words, and the standard deviations of these adjustments varied between 12 and 27 ms.. When the results from all three subjects were combined, the mean durations of vowels in production closely matched those set in perception by the same three subjects. These results are of only limited relevance to the questions posed in the present study, for two reasons. Firstly, Huggins (1972a) has shown that subjects are more sensitive to changes in vowel duration than to changes in consonant duration. Secondly, Rudolph's (1971) stimuli were isolated words, whereas it was the intention in the present investigation to include the stimuli in complete sentences, so as to make the experimental situation more natural and to assess the effect of different speech contexts on the subjects' performance.

Somewhat more pertinent to this study are the results of Huggins (1975 a, b). Huggins (1972a) studied the jnds for durations of /s, /, a, i/ and stressed /a/, using natural speech stimuli. All consonants were in initial unstressed pre-stressed position; /a/ was also presented unpredictably and word-finally. Apart from the differential sensitivity to vowels and consonants mentioned above, Huggins showed changes in segment duration to affect perceived stress and perceived rhythm. The range of durations that subjects accepted as normal for a given segment varied according to the position of the segment in the word, the immediate phonetic context, and the overall stress pattern. In a further series of experiments studying perception of the duration of /a/ and stressed and unstressed vowels only,

Huggins (1972b) confirmed and extended his earlier results. He concluded that perception of time in running speech is more dependent on events at the syllabic level (particularly connected with stressed vowel onsets), than at the segmental level. Huggins also found some slight evidence that subjects were less tolerant of durational changes that lengthen segments than of those that shorten them (1972a, p.1276).

The relevance of immature temporal patterns in speech production to the perception of segmental duration

Consonantal segments in adult speech characteristically undergo particular forms of temporal reorganisation according to the phonetic context in which they occur. For example, /s/ alone, as in 'say', would be expected to be longer than when clustered with other consonants, as in 'spray'. The extent to which the segment would be shorter however depends upon the particular consonant with which it is clustered.

Hawkins (1976) showed that such mature forms of temporal reorganisation are often lacking for some segments of children's speech. The immature temporal patterns can take a number of forms, depending upon the individual child speaker and the particular phonetic sequence. One very common immature pattern is a failure to shorten the duration of consonantal segments in the appropriate contexts; elongations relative to the segments' unclustered durations can also be very frequent.

Although these adult-child differences for consonant duration in production are often highly significant statistically, the main results of Huggins' (1972a, b) studies suggest that some of them are not individually perceptible. Moreover, it is quite possible that they are not relevant to perception at all, even in interaction with other factors. However, the possibility raised by Huggins' (1972a) results that sensitivity may be greater for lengthened rather than shortened segments acquires more significance since most of the interesting differences between adults' and children's productions involved overlong

segments for the children. Moreover, it seems reasonable to suggest that adults' and children's sensitivity to speech duration may differ, and that the differences may reflect differences in underlying processes mediating both perception and production of duration. It is also possible that sensitivity in the perception of segmental duration may depend upon the speaker's age, for either adult or child listeners, especially the latter.

An investigation of both adults' and children's perception of consonant duration in comparison with durations of the same segments in their speech production, then, could contribute towards a greater understanding of the relationship between the regulation of timing in perception and production. It could also indicate some of the causes of the adult-child differences observed in production. They could stem simply from a muscular inability to carry out the necessary motor gestures; or they could be due to an immature central timing regulator transmitting commands for the sequencing of gestures that do not correspond to those of adults.

Summary of the questions asked

1. What are the jnd's for the duration of consonants in clusters? Some general indication of this is necessary before the importance of the observed differences with children can be evaluated. Some preliminary data we have collected indicates that, under the particular, somewhat reduced, conditions of the experiment, jnd's are in fact closer to the 30-40 ms proposed by Lehiste, rather than to Nooteboom's 10 ms.

2. Are the characteristic durations of clustered consonants in the speech of adults also 'preferred' by them in their perception of speech? This is relevant to the question of the generality of the speech timing mechanisms.

3. Do the immature patterns in children's production of consonant clusters stem from muscular constraints or from an immaturity of the central timing regulator?

Hypotheses

A strong connection between the timing regulators for speech production and perception would be implied if adults' preferred perceptual durations corresponded to their preferred production durations. Complete identity of the regulators would in fact be likely if this were the case. If we find that children's performance in production and perception is also similar (but different froms adults'), then we could infer that their non-adult speech patterns stem from a central immaturity. If children's preferred perceptual durations are similar to the adults', though, a motor inadequacy would be indicated as at least the major cause of their atypical production patterns. If children's perception corresponded neither to their own patterns of production, nor to adults' perception, then immaturity of both the central timing regulator (or regulators) and of the speech musculature would be implied.

A nonidentity between preferred consonant durations in adults' perception and production would imply either that the timing mechanisms for production and perception operate at least largely independently of each other, or that the speech apparatus is not capable of carrying out the the timing regulator's commands very exactly. As there seems to be little reason for speech perception to be significantly more precise than production, the latter possibility is somewhat unlikely. The independence hypothesis would be favoured in this case. If children's perceptual performance corresponded to the adults', in this case, one could infer an adequate perceptual system, but a production system which was inadequate either with respect to central commands for timing, and/or with respect to the ability to translate the commands into muscular action. Since the perception of timing is mature, it is perhaps more likely that muscular inadequacies are responsible for at least the major part of the production difficulties. But this would be by no means certain.

If, on the other hand, children's perceptual performance differed from that of the adults (but did not correspond to their own produced durations), an immaturity of the central regulation of both production and perception would be implied.

From the adult evidence, the two aspects would seem still to be independent in these cases, but an added muscular inadequacy would not be ruled out.

Finally, it is possible that children's perceptual performance could correspond to their patterns of production, and yet adults' production and perception could be dissimilar. This would imply that in children, timing production and perception are closely linked, at least for consonants, and that during maturation, the two functions become dissociated. The link might be in the form of a feedback loop utilised during the acquisition of motor speech skills and not used latter. The possibility of motor inadequacy as the sole cause of the observed production anomalies would be rules out.

Choice of design and materials

Despite inherent inadequacies, synthetic rather than real speech was used on-line. This allowed the stimuli to be continuously variable rather than presented to subjects in discrete steps. It also allowed the critical segments to vary over a very wide range of durations without requiring the time-consuming construction of each separate stimulus that would be involved if natural speech stimuli were used. It was thought preferable to sacrifice naturalness to continuous variation of the segmental durations over a wide range, since as much flexibility in design as possible was wanted in order to constrain the child subjects' behaviour as little as possible. In this way it was hoped to tap any adult-child differences that were present.

The method of adjustment to an internal criterion was chosen, largely for considerations concerning child subjects, as outlined above. This method was used successfully in a similar situation (with adult subjects) by Nooteboom (1973), who lists a number of experiments by other people in auditory and visual research that used the same method.

The consonant clusters selected for study were /str/ and /st/, along with unclustered /s/ and /t/. These were chosen because Hawkins' (1976) data had indicated that clusters with initial /s/ and those with stops differ between adults' and

children's speech production in particularly interesting ways, including a high incidence of segmental lengthening in clusters produced by children. The homorganic /st/ sequence was chosen rather than /sp/ or /sk/ because it tended to remain deviant in the children's speech for longer. Children old enough to understand the perception task and to cooperate for long enough would therefore be more likely to show immature patterns for clusters with /st/. The stimuli were to be presented in two carrier phrases, chosen to contrast in overall length, number of words, stress pattern and intonation contour, while at the same time being meaningful and relatively easy to synthesise realistically.

METHOD

Subjects

Three adults completed the experiment. Subject A1 was female, 27 years old. She had not heard synthetic speech before, nor taken part in an experiment of this type. Subjects A2 and A3 were male, 29 and 28 years old respectively. Both are experimental phoneticians, with considerable experience of synthetic speech. A3 in particular has a detailed knowledge of durational parameters of adult English speech. None of the subjects knew the detailed expectations of this experiment, nor had more than general knowledge of the durational characteristics of child speech. A fourth adult, a female, began the experiment but dropped out because it was too time-consuming.

Two children completed the experiment. They were sisters aged 7,6 to 7,10 years and 5,11 to 6,3 years respectively during the experiment. Subject K1 is the older, and K2 the younger. Four other children began the experiment but dropped out for various reasons. One boy and one girl, both about 4,4 years, failed to understand the task sufficiently and refused to continue. (It is possible that they could not yet isolate a single phonemic sound from the sentence as a whole.) One boy and one girl, 6 and 5 years old respectively, performed the task well but left due to time considerations.

All subjects reported normal hearing.

Terminology

The following terms are defined:

word	a critical word heard in either or both carrier phrases e.g. 'say'
stimulus	a combination of one experimental (critical) word in a particular carrier, e.g. 'say' in 'Please do say now'
critical segment	a segment of a word whose duration is varied e.g. /s/ in 'say'

chunk	a group of slices to be synthesised that corresponds to a phonetic segment, or a group of such segments
trial	one setting of the experimental wheel (see 'Procedure'), leading to one synthesis of the stimulus
run	number of trials until subject signals he is satisfied
session	group of uninterrupted runs on the same stimulus

Notation

Throughout the text, the following notation is used:

/eI/ represents the vowel as in 'say'

All segments that were produced at their ideal durations are written in lower case; the critical segment, presented to the subject as either too long or too short, is written in upper case. Thus /SteI/ represents 'stay' with /s/ as the critical segment, and /sTei/ represents 'stay' with /t/ as critical segment. Different stimuli involving the same word (e.g. 'say' in both carriers), are distinguished from each other by adding 'g' or 'n' to the end of the phonetic transcription of each, according to whether the word occurs in 'Now you'll hear --- again', or 'Please do --- now'. respectively. 'Words' are written with the critical segment in upper case and the whole between slashes, as for phonetic transcription. 'Stimuli' are written in single inverted commas: thus 'sTeIg' represents the stimulus 'now you'll hear stay again' with /t/ as critical segment.

Stimuli

The stimuli were synthesised on an Ove II B synthesiser (Fant et al., 1962) using SPEX (Draper, 1972, 1973). SPEX is based on Focal, a language written for PDP 12 computers, and is designed specifically for use in on-line interactive experiments with a PDP 12 and an Ove III.

The words selected were: 'say, tay, stay, stray', (footnote 1). Each was presented in two carrier phrases, also synthesised on the Ove III. These were "Now you'll hear —— again" and "Please do —— now". The stimuli were synthesised in both an adult male voice and a child's voice. Parameters for the male voice stimuli were taken from Peterson and Barney (1952), Peterson and Lehiste (1960), Fant (1968), Hawkins (1976) and Klatt (1973a,b, 1974). Values for the formant and amplitude parameters of the child's voice were taken from Peterson and Barney (1952), Fant (1968) and Simon (1976); durational values were from Hawkins (1976). Spectrograms of one adult male and one six year old boy saying the sentences to be used were also consulted. The final values of all parameters were arrived at by adjusting until the utterances sounded as natural as could be achieved. To make the task simpler for children, the overall durations chosen were toward the long end of the range for production at a 'comfortable' or 'normal' pace.

In order to produce stimuli where the duration of just one phonemic segment would be varied, the utterances were divided into several 'chunks': the carrier phrase preceding the critical word; each of the critical segments in the word; the rest of the word and the remainder of the carrier. All chunks were stored separately but could be synthesised with no pause between them, so that the perception was of a fluent sentence. Durational values for each 'slice' in the critical segments were set at the minimum possible (2 ms), and the actual number of slices was kept as small as possible within the limits of obtaining smooth movements of parameters and thus high quality of synthesis. In this way the critical segment would be far too short if synthesised at the stored rate; but by synthesising it at increasingly large multiples of the stored rate, it could be made progressively longer until it was heard as far too long. It was this rate of synthesis of the critical segment that subjects controlled.

The ideal durations of the critical segments are shown below.

Ideal durations (ms) of critical segments

	Adult Voice	Child Voice
/SeI/	156	204
/SteI/	110	175
/StreI/	104	168
/TeI/	60	72
/seI/	54	85
/sTreI/	54	80

spectrograms of each utterance in each carrier are shown in the figures.

Procedure

The experiment was run on-line. The control program, written in SPEX, initiated a routine whereby the subject effectively indicated the range of durations acceptable to him for a given acoustic segment in a particular phonetic context. The duration of the critical segment was governmed by the position of a 'steering wheel' which was connected via a potentiometer to an analog-to-digital converter of the PDP 12. The digital value received by the computer depended upon the position of the wheel. A clockwise turn increased the value; an anticlockwise turn decreased it. (The maximum amount of turn was just less than 360°). The relationship between the angular wheel position and the value was linear.

The wheel had a round white button in the centre, attached to a horn. This was only to amuse subjects. To prevent position cues being used by subjects, the wheel had no spokes or other markings and was uniformly painted dark green. As a further precaution against the use of position cues, stimuli were synthesised so that their ideal durations required synthesis at a range of different rates. Thus the wheel had to be at a different point in the total turn for each one. The ideal rates for synthesis of each critical segment are shown below.

Ideal rates for synthesis of critical segments (%)

	Adult voice	Child voice
/SeI/	600	600
/SteI/	690	625
/StreI/	650	600
/TeI/	750	900
/sTeI/	670	850
/sTreI/	675	800

As a final precaution against position cues being used, a given wheel position did not always lead to the same rate of synthesis, across runs of trials. Thus the value of the A to D converter was multiplied by 1.00, 1.02, 1.04 or 1.06 before the rate of synthesis was computed from it. This constant was randomly selected at the beginning of each new run of trials, and remained the same for that run. If it had been changed for each trial, subjects' estimations of the amount of turn required to produce a give result would have been disturbed.

The range over which each critical segment could vary in duration was from 90% to 2500% of the stored value for all words except /TeI/, where it was 85% to 4000% of the stored value. The larger range for /TeI/ was necessary due to the shortness of the stimulus, and the fact that the subject was manipulating silence only. Thus the absolute durations over which stimuli could vary differed between stimuli. The ranges of absolute durations for each stimulus are shown below.

Ranges of absolute durations (ms) over which critical stimuli varied

	Adult voice		Child voice	
	min	max	min	max
/SeI/	23.4	650	30.6	850
/SteI/	14.4	400	25.2	700
/StreI/	14.4	400	25.2	700
/TeI/	6.8	320	6.8	320
/sTeI/	7.2	200	9.0	250
/sTreI/	7.2	200	9.0	250

All subjects were tested individually. They sat in front of the experimental wheel in a double-walled sound treated room, and heard the stimuli through high quality head phones. Volume was adjusted to a comfortable level for each subject. The subject heard a sentence in which all segments were at their 'ideal' (normal) durations except the critical one. At the beginning of each run of trials, the duration of this critical segment was either maximally long or maximally short. Starting alternately from the maximum possible duration on the critical segment, or the minimum possible duration, the subject's task was to alter the position of the wheel until the duration of the critical segment in that sentence sound acceptable to him. If the wheel had not been moved for 3 seconds, the value received by the computer from the wheel position was converted to a rate at which the 'slices' of the critical chunk were synthesised in the next stimulus presentation, which followed immediately. Thus for each new position of the wheel, the subject heard the sentence with the duration of the critical segment altered by an amount proportional to the distance he had turned the wheel, all other parameters throughout the sentence remaining unchanged. The subject signalled when he was satisfied with the duration of the critical segment by activating a sense line. This initiated the next phase of the experiment (a new run of trials on that stimulus, or a new stimulus).

Subjects were tested on average about once a week. Adult subjects generally did 6 runs (6 'satisfied' signals) on a given stimulus, and would complete between 3 to 6 different stimuli in one experimental session. Children did 4 runs on each stimulus, and usually completed 4 different stimuli in one session. Towards the end of the total testing period, which lasted several months in most cases, the number of runs on one stimulus was sometimes less than usual, especially if extra data were being collected to compensate for unusually large variability in earlier performances.

Rest pauses were allowed at any time, but were discouraged during a run of trials on one stimulus, and were actively encouraged only between stimuli. Within one experimental session, subjects heard two stimuli in one voice (adult or child) and then two in the other voice, and so on. Usually, the carrier phrase, the critical segment and the

word it occurred in all differed between successive stimuli. Where this was not possible (particularly towards the end of the experiment when data sets were being completed), as many of these parameters as possible varied between successive stimuli.

The experimenter was present during all sessions with the children. She heard the stimuli over head phones at the same time as the subjects did. She interacted with child subjects continuously, asking them questions about each stimulus as it was heard (see 'Instructions') and often turning the wheel for the child in accordance with her (the child's) instructions. This was partly because the children sometimes preferred just to play with the wheel (and were rather unpredictable about when they would start to do this), partly because they enjoyed alternating between turning the wheel themselves and instructing the experimenter to, and partly to encourage the child to attend to the task. The last reason became particularly important towards the end of experimental sessions.

To keep conditions as similar as possible between adults and children, the experimenter was present during all sessions with subject A1, and for all but some of the later ones with A2 and A3. She interacted with the adult subjects only when necessary to encourage them to be more attentive and careful (see 'Instructions') and in casual conversation during rest pauses.

The experimenter's presence with the subjects made it necessary for the teletype to be in the same room also, since instructions for a given stimulus were given at the beginning of the first run for that stimulus. However, no printing occured while the subject heard the stimulus, and the teletype's noise level when it was not printing was satisfactorily low. No subject complained when asked if it was disturbing.

Instructions

Adults were told that the experiment was concerned with the perception of speech sound duration. They were not told any details about the structure of the stimuli, though phoneticians A2 and A3 were of course aware of general factors

involved for adult speech. Whenever it seemed necessary during the course of the experiment, it was stressed that there was no objectively 'right' response. It was also stressed, however, that performance should be reasonably consistent within stimuli. If a subject's variability began to increase beyond his normal level, this was pointed out to him and he was asked to try to be more careful. If the cause seemed to be due to one particular factor, such as moving the wheel in large rather than small steps, this too was mentioned. After the first session, no adult needed to be reminded of the purpose of the experiment or of the task required of them.

At the outset of the experiment, verbal instructions as to the procedure were kept to a minimum as these tended only to confuse subjects. No subject had any difficulty in understanding the task, however, when he began to operate the wheel and adjust the duration of a segment. The segment /s/ in 'say' was always used as the first example, as it was easiest to deal with. Experimental trials began after the subject felt he had grasped the procedure and felt confident with his performance. This usually took two or three runs of trials, of which the first run was usually rather longer than average - about 18 trials instead of the average of about 8. When a stimulus was introduced for the first time, the first run of trials was discarded.

Before beginning each new stimulus, the subject was told the sentence he would hear, in which voice it would be, which sound in the critical word he was to alter, and whether he was to start with the sound maximally long or maximally short.

Children were told that the experiment was to see 'how they thought people ought to speak'. They accepted this as a valid purpose. As with adults, it was stressed at the outset and throughout the experiment that there was no 'right' answer, but that the experiment was interested entirely in 'how you think it should be'. The task was explained with respect to /s/ in 'say' and further instruction continued entirely in conjunction with the child hearing the stimuli and turning the wheel to produce new values of the critical segment. The children soon learned to turn the wheel by an amount they thought appropriate, and then to take their hands away from it until they heard the next stimulus. The

control program contained an option such that, instead of synthesising the utterance if the wheel had not moved for 3 seconds, the program paused indefinitely after each synthesis until it received an instruction to continue. The children were introduced to the procedure and to each new stimulus with this pause option implemented, to allow ample time for consideration and questions between each presentation of the stimulus.

During experimental trials, each child chose whether she would work with the pause option implemented or not: these were termed the 'slow way' and the 'fast way' respectively. Subject 01 usually chose not to implement the pause option, and thus had the same conditions as the adults. Subject 02 varied, utilising the option roughly 10% of the time, usually depending upon how tired and cooperative she felt.

As with adults, the children were told before a new stimulus was presented which sentence they would hear, in which voice, which critical segment they were to alter, and whether to start with the sound 'as long' or 'as short' as possible. They were asked to repeat the sentence and the critical segment to the experimenter before trials began.

A minimum of 12 runs (trials to a 'satisfied' signal) were collected for each stimulus from each adult subject. A minimum of 3 runs were collected from each child subject. The complete data set for each stimulus thus came from at least two sessions.

Speech production

Ten utterances of each word spoken in both carrier phrases were collected from each subject after the last session of the perceptual task. The sentences were randomised in blocks of 13. Adult subjects read them, and child subjects repeated them after the experimenter. All subjects were encouraged to speak as fluently and as naturally as possible, maintaining a constant rate. The sentences were recorded on a Nagra tape recorder playing at 7½ i.p.s. in a double-walled sound treated room.

RESULTS AND DISCUSSION

Preliminary analysis on the variances

The data appeared to be sufficiently variables between subjects, stimuli and experimental sessions to make it necessary to examine the variances of the results in order to choose appropriate statistics for the main analyses. Accordingly, variances for each subject on each stimulus (word x carrier) and each experimental session were plotted on a logarithmic scale (to minimise differences due to absolute duration of the stimuli). Where there were data from more than two experimental sessions for a given stimulus, the last two sessions with a complete number of runs, or one less than the complete number, were taken for this analysis. (Thus sessions with 5 or 6 runs were used from adults, and with 3 or 4 runs from children.) Data for /Te/ were excluded since there were none from the child subjects.

The plots showed the variances to be unimodal and positively skewed, allowing an analysis of variance to be run to test for homogeneity of variances across variables. Experimental sessions were again kept separate for each stimulus. The model thus had 5 factors: words, (divided in terms of critical segments), voices (adult or child), and sessions (first or second) were fixed; subjects and carriers were considered random.

This analysis is summarised in Table 1. Assuming the above model, one tailed tests of significance gave the following significant effects:

words	$p < 0.01$
carriers	$p < 0.05$
words x voices	$p < 0.01$
voices x carriers x subjects	$p < 0.01$
words x voices x subjects x sessions	$p < 0.05$

Assuming a model with all factors fixed, the above F ratios remain significant, with the level for carriers rising to p.0.01. In addition, significant F ratios occur for:

voices	$p < 0.01$
sessions	$p < 0.01$
voices x sessions x subjects	$p < 0.01$

(For arguments for and against the assumption of models with random variables versus fixed effect models, see Clark (1973, 1976), Cohen (1976), Forster and Dickinson (1976), Keppel (1976), Smith (1976), and Wike and Church (1976).)

Given these results, it was considered inappropriate to run a single analysis of variance on the mean data, since differences in the variances occurred between the factors of interest. Instead, a number of separate tests were run on different aspects of the mean data. For these tests, the data were selected as described in the next section, unless otherwise stated.

Data selection for the main analyses

The original aim was to have 12 data points from each adult subject for each stimulus, and 8 data points from each child subject. Extra data was collected, however, if a subject's performance during the second session on a given stimulus was different from his performance in the first session. In these cases, data were taken from the sessions that were most similar to each other in terms of mean duration and range of response. As far as possible, the data chosen were restricted to those from only two sessions. Occasionally, 1 or possibly 2 points in one session lay well outside the range of the rest of the data: in these cases, they were replaced by data randomly selected from another session on the same stimulus. Thus a basic set of 12 points for each adult, and 8 for each child, were selected for each stimulus. Experimental sessions within stimuli were not distinguished in these data sets.

Next, data for 1 word were combined across the two carriers, giving 10 data sets in all, with 24 points in each for adults and 16 for children. These data sets were then converted to logarithms (base 10). The mean of the logs. for each set was found, and each data point was then expressed as a normalised deviation from the mean for the set, using the formula

$$\frac{\log_{10} x_i - \overline{\log_{10} x}}{\text{s.d.}}$$

where x_i = a data point (log), $\bar{x}$ = the log mean of the data set, and s.d. = the standard deviation.

The data for each data set were then inspected, and items that deviated considerably more than others in the set were deleted if that set had a large variance. In all, 14 data points were deleted in this way: no stimulus (word x one carrier) had more than one deletion per subject. Table 2 lists the deletions made, with carriers separated. This basic set of data was used in all subsequent analyses (carriers separated).

Correspondence between durations in production and those set by perceivers

The perception data were compared with data for production of the same words and sentences by the same subjects, and also with the same consonant clusters in the data of Hawkins (1976). Hawkins' data are for 5 adults and two partially overlapping sets of 7 children. The age range in one set, KY1, is 4 to 7 years, and in the other set, KY2, it is 4 to 8 years. (6 of the 7 children appear in both sets, being one year older in KY2.) Hawkins (1976) used different words from those in the present experiment and her subjects said the words in isolation except for a preceding /ə/. Therefore the most phonetically similar words are used for comparison. Those for the 5 adults and the children of KY2 are /sed/, /tel/, /step/ and /strɪŋ/. For KY1, some words involving the same vowel in CV or CVC monosyllables (with the final C varying in manner and/or place) were pooled. The words used in this comparison are /saI(C)/ and /teI(C)/. Words used from KY1 without pooling over final segments are /step/ and /strɪŋ/. The mean durations (ms) set by each subject in the perception task are shown, with the standard deviations, for each stimulus in Tables 3 to 6. These tables also show the mean durations (ms) and standard deviations of the subjects' production of the relevant segments, and the range of mean durations for subjects in Hawkins' (1976) study (for segments in the most phonetically similar words).

It is unreasonable to expect exact correspondences between the produced and perceived durations of the same subject. The parameters involved are very complex and the synthetic speech was not modelled exactly on each subject's individual productions. Comparisons with Hawkins' (1976) data are even more indirect, since the subjects contributing to the two sets of data are different and the speech contexts differ. Moreover, a relatively slow overall speech rate was used for the synthetic stimuli, whereas subjects were asked to speak at a comfortable rate, but not slowly (see Method). This is reflected in the general tendency for each subject's production durations to be less than the values he set the same segments to in the perception task, especially for /s/. (A3 is an exception to this tendency.) In general, all the production durations of subjects A1, A2, A3, K1 and K2 fall within the range of the Hawkins (1976) means, whereas some of the durations set in the perception task exceed the production durations quite considerably. However, Table 7 shows that the Spearman rank correlation between individual subjects' production and perception durations of the same sentence is high. $r_s \geq 0.85$ ($p < 0.01$) for

all subjects except K1, for whom $r_s = 0.66$ ($p < 0.05$). (The objection could be raised that these high values are partly forced due to differences in the absolute durations of /s/ and /t/. Tables 3 to 6 show that ranks within segment type (/s/ or /t/) are still very similar between production and perception for each subject. In this case therefore, the forcing effect is minimal.

Given this intra-subject consistency, let us consider general patterns across all subjects for each set of stimuli. General measures of consistency were assessed by computing Spearman rank correlations (r_s) if only two sets of data were being compared, and Kendall coefficients of concordance (W) with more than two sets of data. These two measures are linear with respect to each other (Siegel, 1956; p.232). The following comparisons were made: the three adults on durations in perception only, in production only, and in both production and perception together; the two children on durations in the same three sets of data; and all five subjects together for the three sets of data.

Values for W or r_s, as appropriate, are shown in Table 8. All the coefficients are significant, most at beyond the 0.001 level of probability. The least degree of association is between K1 and K2 on durations set in the perception task, but even this is significant at $p < 0.05$, with $r_s = 0.694$. Note also that most of the values for measures between subjects in Table 8 are rather higher than the values for intra-subject comparisons in Table 7. This implies that while there is a significant association between the same subject's behaviour in the perception and production of the segments in these sentences, there tends to be a rather stronger similarity between different subjects performing the same tasks. None of these differences are significant, however.

We can look for differences not shown up by the correlational analysis by returning to Tables 3 to 6. Comparing Tables 3 and 4 with Tables 5 and 6, it can be seen that there are rather more differences among all subjects between the perception and production data for the child's voice than there are for the adult's voice. To illustrate this point, the average difference (in either direction) between the same subject's perception and production durations is 32.2 ms for adults and 62.5 ms for children. (The difference for adults would be smaller were it not for A2's long /s/ durations in the perception task (Table 3).) The values set in perception by the children responding to the adult voice, and by the adults responding to the child voice, maintain this larger difference between perception and production durations for the child voice.

The voice-dependent effects are rather more exaggerated in the /st/ cluster (but /t/ in /st/ with the adult voice stimuli also tended to be set rather long). A2 has a general tendency to set longer durations than A1 or A3, and often longer than K1 and K2 also. K1 and K2 have a relatively marked tendency to lengthen /s/ in the child voice beyond the overall production range more than they did for the adult voice stimuli. This is particularly interesting, given that the durations of /s/ in K1's and K2's speech production tend towards the shorter end of the range of /s/ durations produced by the adult subjects. This is discussed in more detail later.

In general, then, the durations set in the perception task are sufficiently close to those obtained from the production data to allow the conclusion that subjects were able to manipulate the synthetic stimuli until relatively realistic durations were achieved. With the exceptions of A2 and of the child subjects hearing the child's voice, the differences between the two sets of data can be explained in terms of differences in speech context and overall rate, and the inherent difficulty of the perception task.

Differences between carriers

From the raw data it seemed that some subjects were considerably influenced in their performance by the carrier in which the word was presented. The variances have already been shown to be significantly different from each other, with that for 'Now you'll hear — again' being about 1.5 times larger than that for 'Please do — now'. To examine the differences between the means for carriers while controlling for differences in the absolute duration of the stimuli, logarithmic data were again used. This transformation was appropriate as there was a general tendency across subjects for larger variances to be associated with larger means. Each observation was further transformed into its deviation from the mean of the logarithms of items in that data set (two carriers combined), so as to remove as much as possible any nonhomogeneities due to words and subjects. Data for /TeI/ were again excluded since there were none for K1 and K2.

An analysis of variance was computed on these deviations from the mean logs, extracting the factors of voices, words,

carriers and subjects. The model thus had four factors, with repeated measures on all but 'subjects'. All factors were treated as fixed in the estimation of expected mean squares, since the main interest was in carriers and interactions with carriers, and the preliminary analysis of the variances had shown the carriers to be significantly different from each other. It was reasoned that if their variances differed, then other aspects might also differ between them. In this case it was more realistic to treat them as fixed variables, agreeing with Gladstone and Siegenthaler (1971) that it cannot always be assumed that subjects will perform identically with the same word presented in different sentences. Moreover, the two carriers had originally been chosen partly because they represent different stress and intonation patterns; the criteria for treating variables as fixed effects include cases where items have been selected in such a nonrandom manner (Winer, 1971; p.143).

F ratios for the main effects of voices, words and subjects are expected to approach zero, since the data were transformed so that all the data sets had a mean of zero with both carriers combined within voice. However, this procedure meant that a significant effect for carriers could occur, and interactions with carriers could also be significant and of interest. Table 9 shows the expected low F ratios for voices, words and subjects and their interactions. In contrast, the main effect for carriers is large and is significant at well beyond p 0.001. Critical segments in words presented in the context 'Please do -- now' were judged as requiring shorter durations than segments in 'Now you'll hear --- again'.

Some of the significant interactions deserve consideration, but most are uninterpretable. The transformation of the data into deviations from the logarithmic mean prevents them being interpreted in terms of milliseconds. Values of these interactions are thus not useful to the reader, so they are only presented for the two most interesting interactions (see below). The following interactions, although highly significant statistically, (see Table 9), cannot usefully be interpreted: words x carriers, carriers x subjects, words x carriers x subjects, voices x words x carriers, and voices x words x carriers x subjects. The words x carriers interaction depends upon a reversal for /sTr/ of the general rule that durations are shorter with 'Please do --- now' than with 'Now you'll hear --- again'. Similarly, carriers x subjects depends upon subject AI reversing the same finding. The interaction of words x

carriers x subjects, voices x words x carriers and the four-way interaction show the above effects but involve too many other inconsistencies to make it worthwhile describing them.

The interactions of voices x carriers, and of voices x carriers x subjects are rather more interesting. Table 10 shows that the adult voice led to shorter duration judgements in the context of 'Now you'll hear --- again' than in the context of 'Please do --- now', contrary to the overall direction of difference between the two carriers. In contrast, the child's voice induced shorter durations with 'Please do --- now'. The difference between carriers is greater for the child's voice, thus causing the direction of the main effect difference. These differences are probably stimulus-specific and therefore of little general interest without further research into the role of different pitch, stress and intonation patterns on consonant segment durations.

The three-way interaction, voices x carriers x subjects, promises to have more general interest, however. Table 11 shows that with the adult voice stimuli, adult subjects set durations longer with 'Please do --- now' than with 'Now you'll hear --- again'; whereas they set shorter durations with 'Please do --- now' in the child's voice, except for A1. The children, on the other hand, set durations with 'Now you'll hear --- again' that were longer than with 'Please do --- now' for both voices. The voices x carriers interaction may thus be completely contained within voices x carriers x subjects, the effect of the three adult subjects outweighing that of the two child subjects. Given the exception of A1, it cannot be concluded from these data that this adult-child difference will always be found, but it is possible that the children are exhibiting somewhat more rigid behaviour than the adults. If this is the case, then it is likely that the children are not influenced as much as adults by general linguistic and phonetic factors of the speech context.

These differences dependent upon carrier phrase and voice, then, suggest that further research into the effect on duration perception in speech contexts with varying characteristics would be of interest in its own right. Furthermore, different linguistic contexts may influence adults' more than children's perception of speech durations: these differences could provide insights into the development of mature perception of speech duration, and possibly into mature production of temporal parameters in speech development. The next section considers the effect of stimulus voice on adult versus child subjects in more detail.

Differences between adult and child perceivers in terms of stimulus voice

Mann-Whitney U tests were performed for adult versus child listeners on each stimulus (word x carrier separately), excluding /Tel/. The results are shown in Table 12 for adult voice stimuli, and in Table 13 for child voice stimuli. From these tables it can be seen that of the 10 comparisons per voice, adults and child listeners differ from each other significantly on 4 adult voice stimuli, and on 8 child voice stimuli. In each of the adult voice cases, the children set shorter durations than the adults did; in contrast, the children set all 8 of the child voice stimulus durations longer than the adults did. This consistent difference of direction lends greater importance to the 40% of the adult voice stimuli that showed adult-child listener differences.

The children thus appear to be behaving differentially with respect to adult and child voices in a way that adult listeners are not. The nature of the difference suggests that the children are over-compensating in perception for a difference that occurs naturally in production - that is, they "expect" the child's speech to be slower than the adult's, but over-estimate the extent of the difference compared with adult listeners. These points will be returned to in the final discussion.

Is there an 'internal criterion' for segment duration?

The first stated aim of this research was to determine the just noticeable differences (jnd's) for durations of consonants in clusters. Normally, these should be clear from the standard deviations of subjects' responses. In this experiment however, subjects had relatively wide freedom of response compared with most psychophysical experiments on jnd's, for reasons discussed elsewhere (see Criticism of the Method). The rather high response variability engendered by the design, coupled with the need for data for one stimulus to be collected over more than one session produced an 'artifact' of both practical and theoretical importance. That is, it was often the case that, for a given stimulus, a subject would perform with relatively small variability in both sessions, but that the mean durations for the two sessions would differ quite considerably. Hence the total variance for the two sessions combined would be larger than for either session taken alone.

This effect was not found for all data sets, but all subjects showed some cases. It is distributed randomly across all stimuli. It does not appear to depend upon the day of testing, as it is distributed randomly across all records for each subject. Nor does it appear to be related to particular strategies used by the subject, such as moving the wheel in large steps and gradually 'centering in' on the final accepted duration, versus moving it in small steps in one direction only. The effect was one of the reasons why extra data were collected for some stimuli with some subjects, as described earlier. (The other reason was the presence of unusually large variability in previous sessions.)

To assess the extent of the effect, within-session variances were compared with between-session variances for all stimuli except those with /Tel/. All experimental sessions with 4 or more data points were used for this analysis, rather than the two sets per stimulus used in the previous analyses. Within-session variance was computed by averageing the variances of all 217 sessions; this value was 488.52 ms^2. Between-session variances were calculated in the following way. The differences between the means of all possible pairs of sessions for each stimulus were squared. There were 143 such differences in all. They were summed and divided by 143-1 to obtain the variance. That is,

$$s^2_{\bar{x}_1-\bar{x}_2} = \frac{\sum_{i}^{k} (\bar{x}_{i_1} - \bar{x}_{i_2})^2}{k-1}$$

where $s^2_{\bar{x}_1-\bar{x}_2}$ = sample variance of the means of two different sessions on the same stimulus ($\bar{x}_{i_1}$ and $\bar{x}_{i_2}$), k = the number of pairs of sessions from which the differences were computed.

Substituting,

$$s^2_{\bar{x}_1-\bar{x}_2} = \frac{162723}{143-1} = 1145.937 \text{ ms}^2$$

This value was then transformed as follows: assuming that $\bar{x}_1$ and $\bar{x}_2$ are from the same normal distribution with mean = μ and standard deviation = σ, then

$$s^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma^2}{(\frac{N}{2})}$$

where N = the average number of data points in each session. Thus the between-session variance is

$$s^2_{\bar{x}_1-\bar{x}_2} \times \frac{N}{2}$$

Substituting,

$$s^2_{\bar{x}_1-\bar{x}_2} \times \frac{N}{2} = 1145.987 \times \frac{5.37}{2}$$
$$= 3075.7 \text{ ms}^2$$

An F statistic was then computed on the two variances. The degrees of freedom for within-sessions variance equals the total number of sessions multiplied by the average number of runs per session minus one; this value is 948. The degrees of freedom for between-sessions variance equals the total number of differences

between sessions, which is 143. Thus

$$F = \frac{\text{between sessions}}{\text{within sessions}}$$

$$F = \frac{3075.7}{488.52}$$

$$F_{143,948} = 6.296$$

With a critical F of 1.32 at $p = 0.01$ (df 120, 00), this finding is highly significant. It means that in adjustment situations like the present one, subjects can select an internal criterion of duration acceptability and adhere to it relatively closely over successive repetitions within one session, but that their criterion can drift quite considerably between sessions separated by days or weeks.

It could be argued that this difference is an artifact arising from the inherent lack of rigour of the experimental design. While the phenomenon would certainly be encouraged by the design, there are strong arguments that it is not a simple artifact that should be ignored. The average within-sessions standard deviation for all stimuli is $\sqrt{488.52} = 22.1$ ms. This value compares favourably with the 25 ms. standard deviation found by Nooteboom (1973, experiment II), whose subjects adjusted the durations of synthetic vowels in words, data for each stimulus being collected in one session. The data for within-session behaviour thus appear to be replicable. We must therefore consider the drift in criterion between sessions in more detail.

Nooteboom (1973) did not present data on judgements from more than one session with the same stimulus (footnote 2). However, Huggins (1972b) reported differences between means for the same stimuli in different sessions. His data were of a rather different type: he lengthened or shortened the duration of either /p/ or a vowel (stressed or unstressed) in a naturally spoken sentence. A segment adjacent to the altered one (/p/ in the case of vowels, and a vowel in the case of /p/) was then lengthened or shortened by progressively larger amounts. Subjects were presented with the pre-recorded sentences and asked to judge whether the segments sounded normal or not in that sentence. His experiment thus controlled the durations of

the stimuli subjects heard, in contrast with Nooteboom (1973) and the present experiment. Data for each stimulus were collected from two sessions on different days. The largest mean differences between sessions found by Huggins averaged 5 or 6 ms. and occurred when the critical segments (CV or VC) crossed word or syllable boundaries; judgements on critical segments within stressed syllables showed differences between sessions of only 1.5 to 2.2 ms, that were not significant with 't' tests.

It seems then, that drift in the internal criterion for judgements of duration is a real phenomenon that must be accounted for and is independent of the discrimination threshold, which remains relatively constant. The size of drift that is measured depends upon the experimental situation: within the normal experimental session it is relatively unimportant; across sessions separated by days or weeks, it can be quite large. If stimuli presented to subjects are from a completely controlled set and are identical in each session, as in Huggins' (1972b) experiment, the drift in criterion will be relatively small but may still achieve statistical significance with some speech materials. (It should be remembered though that Huggins' manipulation of his experimental material was somewhat different from in the present experiment, and this too may have affected the size of the criterion drift.) If the stimulus values subjects hear are not identical across sessions, as in the present experiment, the drift may be much larger. It is likely that a free situation like the present one encourages greater amounts of drift. (The mean difference between different sessions on the same stimulus was 23.35 ms.) Further research is needed to determine the range of drift that will occur across different time periods, with stimuli varying in factors such as overall stress pattern, syllable placement of the critical segments, and with natural versus synthetic speech.

What implications does the drift in criterion over time have then? The most immediate ones affect the definition of a jnd for speech sound duration and the design of experiments investigating these. Rather different are the questions of what process the drift should be attributed to, and what role the drift should play in theories of speech timing and rhythm.

Any traditional psychophysical experiment leaves unresolved the question as to what a subject can do (under optimal conditions) and what he generally does do in 'real life'. The method of adjustment,

allowing greater variability than some other psychophysical techniques and thus tending to produce somewhat larger estimates of jnd, is perhaps a better estimator of 'real life' behaviour than some of those techniques that produce low variances. Nevertheless, it is clear from the present data that the definition of jnd for speech duration still needs to be clarified. The traditional interest, implicit in the term 'just noticeable difference', is in what subjects can do in optimal circumstances. The best definition for speech duration therefore, will be one that examines the threshold at one time. It may be impossible for practical reasons to keep discrimination experiments that involve many different stimuli to one session. In these cases, it seems advisable to continue to collect data for one stimulus from just one session. If comparisons are to be made between stimuli when it is suspected or known that criteria may have drifted appreciably, then the data should be transformed mathematically to minimise the effect of the drift.

The question of what causes the drift in criterion cannot be answered until we know more about what form the drift takes - the maximum amount that can be induced, and whether it fluctuates randomly or in some specific pattern with time. It seems likely that slow changes in body rhythms, affecting such factors as temperature and frequency of electrical discharge from the brain, may prove to be responsible by changing subjective estimates of duration in general. The first few stimuli that the subject hears in the experimental situation may also prove influential. The records of this experiment did not show any sufficiently consistent effects to support this idea, but a controlled investigation of the point is necessary.

The role that drift in criteria for duration should play in theories of speech timing cannot be fully defined until its causes have been at least partially discovered. If it is induced by preceding auditory material, it would seem to warrant a different place in explaining perceptual behaviour than if it is an effect of generalised body states. If caused by the latter, and found to be generalisable to languages other than English, then there are major implications for the relationship between timing in speech production and in perception, and for the nature of the timing regulator for speech.

Finally, it should be noted that these large between-session differences indicating criterion drift add further weight to the evidence that the perception of rhythm in speech depends in large

measure upon the rather precise timing of certain events (largely stressed vowel onsets), while the timing of other acoustic events can vary more widely (Allen (1972a and b), Huggins (1972a and b)). That is, there are certain critical 'time markers' in speech whose onset timing must be fairly precise to sound natural; the durations of segments that are not 'time markers' are perceived relatively less accurately. Hence drifts in criterion for the durations of these segments may be frequent and large. The evidence of Huggins (1972a and b) suggests that consonants before stressed vowels fall into the 'less precise' category. In running speech, these sounds appear to be perceived in a 'temporal slur' that has comparatively broad limits, allowing variation in absolute duration of the speech segments concerned to be rather larger than that allowable for segments that are more critical for the perception of normal rhythmicity. Evidence from Allen's (1972a and b) experiments suggests that the distinction is less clearcut, however, and is not simply one of segments or syllables that are critical for rhythm and those that are not:

> "Although initial consonant release and vowel onset appear to be the major determinants of beat location, however, a rule for precise production of beat location must also include information concerning the articulatory structure of the initial consonant sequence, the degree to which the syllable is stressed, and tempo .... the quality of the nuclear vowel and the rhythmic structure of the utterance must be taken into account as well." (1972b, p.190).

The development of temporal regulation

The second stated aim of this research was to determine whether the characteristic durations of clustered consonants in the speech of adults are also 'preferred' by them in their perception. This question was posed in the context of determining the generality of the regulation of the timing of speech. Thus this question is connected with the third aim of the research, which was to try to deduce whether the immature patterns of timing found in children's speech stem from immaturity in a central timing regulator, or from more peripheral muscular constraints. It has been shown that there is a fairly strong similarity between perception and production for all subjects (Tables 3 to 8), but there are nevertheless too many differences between the perception and production data to allow the conclusion that there is complete identity of the two processes with respect to timing. More research is needed to determine whether the differences found are genuine differences between the regulation of segmental timing in production versus its perception, or whether they stem from other factors. Obvious artifacts of the current situation are the use of synthetic speech, with its inherent inadequacies, and the use of speech parameters which were not the individual subject's own. Ideally, one would use a subject's own speech modified in appropriate ways; the considerable practical difficulties that this procedure involves ruled this method out in a project with the resources of the present one.

Despite these interpretational difficulties, there is enough evidence to conclude that while in many respects children's perceptual behaviour at this age (5 to 7 years) corresponds fairly well with adults', there are some important differences. Some of these differences may be glossed as 'linguistic' or 'sociolinguistic' and others as phonetic. It is reasonable to assume that clear lines between these categories cannot always be drawn however. It has already been noted that the data on differences between carriers points to a lack of sensitivity in the children to general linguistic or phonetic parameters. The precise nature of the parameters has yet to be determined: stress, fundamental frequency contour and small variations in duration are likely to be involved. It appears that children of 5 to 7 years may not be sensitive to some of these factors.

In contrast to this lack of sensitivity, the children seem to be

sensitive to the parameter of speaker voice in a way that adults are not. Specifically, the child speakers distinguished between adult male and child voices by setting relatively longer durations for the critical segments when uttered in the child's voice. They also had a tendency to set relatively short durations with the adult voice stimuli, although this was rather less marked. It was suggested earlier that the children were overcompensating in perception for a naturally occuring difference in production. This explanation is of some interest as a 'sociolinguistically conditioned' phenomenon. More interesting in the present context, however, is the pattern of results within this effect, and its relationship with the production data.

In the children's speech, /s/ was produced with relatively short durations compared with the general relationship between durations in the adults' and children's speech (Table 3 versus Table 5), while /t/ had relatively long durations (Table 4 versus Table 6). This effect has been found by Hawkins (1976) for other children and other clusters involving initial /s/ followed by stops. It implies a different type of organisation of clusters of this type from that found in adult speech. Hawkins (1976) suggested that it represents a stage in development when the child is beginning to integrate the gestures necessary for each segment of the cluster into a unit. She hypothesised that in difficult clusters of this type involving a stop as a non-initial member, the child will pass through a period when he can integrate some but not all of the complex of gestures required for the cluster.

> "In effect, he divides the sequence into two parts. Using /spr/ as an example, he executes the /s/ and the transition into the /p/ closure as a well-integrated complex gesture, rather than as a series of discrete but more simple gestures. He then executes the transition from the /p/ and its release into the following /r/ as a second integrated complex gesture. The integration of the first gesture, however, (/sp/), initially involves the concentration of most of the child's 'planning resources', so that anticipation in terms of preparation for the succeeding gestures is possibly quite minimal. During the stop closure, these later gestures are "replanned" and then executed. This is facilitated by the maintenance of the closure being

> relatively easy, requiring few 'resources' since target overshoot of articulatory configurations is possible. But "replanning" takes time, and leads to an abnormally long closure period. .....
> In the later part of this developmental stage, (when the child has not only attempted, but also achieved, the greater integration of the first two gestures, and of the second and third gestures), the segments surrounding the stop closure should be executed relatively more maturely than the stop itself. They should also be executed more maturely than the corresponding segments in clusters to which "replanning" does not apply." (1976, p.412-413).

The production data from K1 and K2 appear to follow to some extent Hawkins' (1976) notion of "replanning", which was postulated only tentatively. More important in the present context, however, is the fact that the perception data do not seem to mirror the production data in this respect. That is, K1 and K2 set long durations for /s/ as well as for /t/ in the child's voice. This is in contrast to their settings of /s/ with the adult voice, and to their own production values. (Note that there is a weak trend for longer clusters to show greater adult-child differences in duration setting (Table 13)). This may be taken as possible evidence for the unconscious status of "replanning" as a 'strategy' during development, and hence for the relationship between a "replanning" period in the stop and the psychological refractory period (Creamer, 1963; Davis, 1957). (This limits the rate at which successive stimuli, here the internally generated commands for speech production can be responded to.) The data could also be taken as evidence for the independence of timing rules for production and perception at a relatively high level. This interpretation is dubious, however, since it assumes that the children were trying to mimic their own production exactly. This seems unlikely as both subjects thought of the child speaker as a boy but were themselves girls. Moreover, Hawkins' (1976) data showed that some children do not appear to exhibit "replanning", and that some do in fact lengthen both the /s/ and the /t/ in /str/ clusters. It is thus possible that the children in this study were setting durations for /s/ that would be generally acceptable to all children and produced by some, but were not necessarily their own. Furthermore, it is possible that K1 and K2 were setting durations in the perception task

that were applicable to their own speech at a time when they had not yet begun to exhibit "replanning".

In conclusion then, it can be said that this experiment confirms the evidence of Huggins (1972b) and Nooteboom (1973) that the just noticeable difference for speech segment duration is of the order of 25 ms. The experimental method and the materials used were not ideal for determining such thresholds and the way in which they might vary according to the phonetic characteristics of the speech. The method did show, however, that the internal criterion for segmental duration can vary over time by more than the jnd at any one time. This has implications of both practical and theoretical importance. There appears to be quite a strong association between perception and production of segmental duration for both adults and children, but there remain some differences which must be accounted for. The data do not allow conclusions about the generality of the regulator for segmental timing in perception and production, although given the high correlations between individual subjects' perception and production, it seems likely that the two are mediated by a common central regulator but subject to modification during their realisation. Children and adults vary in their sensitivity to various parameters in the speech signal. Further research is needed to clarify the nature of these factors and the causes of the adult-child differences. The nature of the differences observed implies that children lack sensitivity to sophisticated details of the signal that may mediate nuances of meaning, while they are more sensitive than adults to the speaker's age (and possibly sex?). It is possible that this enables them to allow for and thus interpret quite severe variations in the realisation of the 'same' segments dependent upon differences in the speaker's vocal tract size and articulatory ability. It is an interesting speculation that this faculty, if present for parameters other than duration, may account for the frequency with which children are reported to understand each other better than adults can understand them. It may also be important in explaining why children exposed to a foreign language before puberty and early adolescence often acquire that language to a native or near-native level with remarkable facility compared with older people.

Criticism of the Method

The method of adjustment was chosen despite there being some disadvantages inherent to the technique. One major disadvantage is that the technique does not allow for the detection of and differentiation between sequential effects that influence subjects' sensitivity and their choice of criterion. Moreover, it tends to produce flatter response distributions, with larger variances than some other methods. This is because subjects tend to make an initially over-long stimulus gradually shorter by small amounts, and an initially over-short one gradually longer by small amounts; whichever the direction of adjustment, the subject often stops at the first point where the stimulus sounds acceptable. Consequently, the distribution of final values is often artifactually flattened, with too few values in the center of the range, and too many at the extremes. This bias has been discussed by Smith (1957).

In the current context, the method was nevertheless considered to be more appropriate than more precise psychophysical techniques. It was expected to be particularly attractive to children. It was hoped that the inherent freedom of response, while encouraging higher variances, would highlight major adult-child differences in performance. At the same time it avoided the possibility that if, for instance, stimulus step size was predetermined, the values selected would be inappropriate for some or all small children. The exploratory nature of this investigation argued against such built-in constraints, even if it involved sacrificing lower variability.

A further reason for the selection of this method arose during pilot work with the stimuli. Some subjects reported that a stimulus could change from sounding unacceptable to sounding acceptable relatively quickly, but that if they continued to change the duration of the critical segment further in the same direction, the perception would often suddenly become considerably more acceptable, though in no clearly describable way. This phenomenon has been documented in other cases: Cartwright (1941) for instance, discusses it in the context of differentiation between regions of the 'phenomenal field'. A more rigorous standard technique, such as a sequential method, would not reach this 'highly acceptable' portion of the range of acceptability, as with this method, stimuli are concentrated around the thresholds of acceptability at the extremes of the range. By encouraging subjects to 'check' their judgements, it was hoped that final values

judged acceptable might fall within the 'highly acceptable' range of durations. Subjects did not always do this however, since it took much longer and was correspondingly more tiring and frustrating. (Subject A2 was particularly uncooperative in this respect, while A3 was especially careful. To a large extent, this difference in their behaviour is reflected in their results.)

Reasons for the asymmetry of wheel position with respect to ideal stimulus duration

The ideal durations of the critical segments did not fall in the middle of the total range of durations at which they could be synthesised, so they did not fall around the middle of the total turn of the wheel. This rather undesirable situation was allowed for practical reasons. Firstly, SPEX will not reliably synthesise utterances at less than about 81% of the stored slice durations. With very short utterances, values of less than 85% in complex commands would occasionally lead to a breakdown during synthesis. Secondly, the stored critical segments had to contain a certain minimum number of slices so that when synthesised to produce a normal or too long duration, the sound would also be convincingly realistic. These two factors governed the minimum duration of each critical segment. Thirdly, in order to make the task easy for children to understand, and partly to constrain their judgements as little as possible, the maximum rate of synthesis was set at a value that produced a sound that was clearly much too long even to children. In practice, it did not seem that subjects used position cues of the wheel very much in their judgements. They certainly learned that the first turn of the wheel could be large compared with later ones, and the adults (but not the children) commented that the ideal positions were biased towards the shorter end of the continuum. But there are enough cases of subjects moving the wheel across the ideal position far into one extreme or the other, (without stopping to hear a synthesis from near the ideal position), and then returning to an intermediate (ideal) position, to suggest that they were at most only partly aware of the wheel's physical position by the end of a run of trials. Furthermore, the subjects' variances are often relatively large: if position cues were being used effectively, this would not have happened.

Table 1. Analysis of variance for the variances of all stimuli

Source	Sum of squares	df	Mean square	F(1)
voices (V)	1140206.4	1	1140206.44	5.03"
words (W)	4834039.7	4	1208509.92	4.99"
V x W	3043774.0	4	760943.50	4.54"
carriers (C)	1068310.6	1	1068310.63	9.83
V x C	229152.0	1	229152.04	0.33
W x C	261202.5	4	65300.62	0.29
V x W x C	123182.1	4	30795.52	0.12
session (S)	1771615.1	1	1771615.11	2.37"
V x S	216819.0	1	216819.03	0.51"
W x S	168069.6	4	42017.39	0.63"
V x W x S	392507.9	4	98126.97	0.69"
C x S	531102.6	1	531102.60	1.48
V x C x S	30305.0	1	30304.96	0.12
W x C x S	399279.2	4	99819.80	0.53
V x W x C x S	83328.9	4	20832.22	0.17
subjects (Sub)	2865428.9	4	716357.21	1.50
V x Sub	540568.7	4	135142.19	0.20
W x Sub	3563745.9	16	222734.12	0.98
V x W x Sub	3195591.0	16	199724.44	0.70
C x Sub	434716.6	4	108679.15	0.88
V x C x Sub	2763421.4	4	690855.35	5.60
W x C x Sub	3636434.8	16	227277.18	1.84
V x W x C x Sub	4586136.5	16	286633.53	2.32
S x Sub	1479863.6	4	369965.91	1.03
V x S x Sub	3944132.9	4	986033.23	3.32
W x S x Sub	4281758.7	16	267609.92	1.42
V x W x S x Sub	4851726.9	16	303232.93	2.46
C x S x Sub	1437962.8	4	359490.71	2.92
V x C x S x Sub	1186230.1	4	296557.54	2.41
W x C x S x Sub	3020358.5	16	188772.41	1.53
V x W x C x S x Sub	1973099.6	16	123318.72	

(1) Ratios computed for a model with 3 fixed factors (V, W, S) and 2 random factors (C, Sub). F ratios for a model with all factors fixed would have V x W x C x S x Sub as the denominator for all effects.

" F" ratios

Table 2. Items deleted from the basic data sets

Subject	Stimulus	Voice
A1	Strein	child
A2	Seig	adult
A2	Steig	adult
A2	Stein	adult
A3	Seig	child
A3	Streig	child
K1	Sein	adult
K1	Steig	adult
K1	sTeig	child
K2	sTrein	adult
K2	Steig	child
K2	sTeig	child
K2	Streig	child
K2	sTrein	child

1 data point was deleted from each of the above sets. The final number of data points in each of these sets was thus 11 for the adult subjects and 7 for the child subjects.

Table 3. Mean durations and standard deviations (ms) in perception and production for /s/ in an adult voice. Adult subjects: upper row = perception; lower row = production. Child subjects: perception data only. Hawkins (1976): production data only (5 adult speakers); see text for further explanation.

	Word and Context											
	Selg		Seln		Steig		Stein		Streig		Strein	
	m	sd	m	sd	m	sd	m	sd	m	sd	m	sd
Subject												
A1	205	19.35	179	14.69	141	13.54	147	11.74	148	13.89	128	16.93
	[illegible]	6.58	149	6.26	122	14.18	128	7.15	119	6.99	132	9.44
A2	248	29.91	252	18.28	163	16.04	142	8.76	159	20.32	155	21.36
	125	12.47	134	14.87	103	18.30	107	16.34	102	7.15	110	19.29
A3	191	23.22	185	23.41	135	9.21	140	20.40	140	14.34	116	7.11
	178	8.82	185	10.27	154	15.78	163	23.60	150	9.43	161	14.68
K1	180	21.24	176	13.65	147	13.02	157	10.41	113	8.96	121	15.29
K2	170	17.73	175	19.98	125	9.46	140	9.06	163	18.67	164	21.60
Hawkins (1976) Range of means	136 - 201				81 - 178				99 - 163			

In perception data, sample size for adults = 11 or 12, and for children = 7 or 8 (see Table 2). In production data, sample size = 10 for A1, A2, A3, K1 and K2. It is 24 for each adult in Hawkins (1976).

Table 4. Mean durations and standard deviations (ms) in perception and production for /t/ in an adult voice. Adult subjects: upper row = perception: lower row = production. Child subjects: perception data only. Hawkins (1976): production data only (5 adult speakers); see text for further explanation.

	Word and Context											
	Teig		Tein		sTeig		sTein		sTreig		sTrein	
Subject	m	sd	m	sd	m	sd	m	sd	m	sd	m	sd
A1	91	18.13	117	10.96	71	11.54	100	14.01	86	11.83	86	15.39
	86	5.68	81	8.32	62	9.19	56	10.92	51	5.68	48	6.75
A2	109	11.27	89	7.81	77	13.46	99	13.04	70	25.43	123	9.71
	88	7.89	94	18.68	62	5.87	63	10.59	51	9.37	61	13.43
A3	47	8.57	53	12.93	71	15.07	79	14.05		9.25	53	12.04
	95	9.43	89	8.52	55	6.24	60	4.38	49	11.97	51	10.92
K1	-	-	-	-	98	22.93	78	8.20	81	6.34	79	6.61
K2	-	-	-	-	81	18.65	54	4.71	66	8.23	61	8.31
Hawkins (1976) range of means	49 - 96				50 - 70				44 - 89			

In perception data, sample size for adults = 11 or 12, and for children = 7 or 8 (see Table 2). In production data, sample size = 10 for A1, A2, A3, K1 and K2. It is 24 for each adult in Hawkins (1976).

Table 5. Mean durations and standard deviations (ms) in perception and production for /s/ in a child's voice. Adult subjects: perception data only. Child subjects: upper row = perception; lower row = production. Hawkins (1976), production data only (KY1 7 child speakers; KY2 7 child speakers); see text for further explanation.

	Word and Context											
	Selg		Seln		Stelg		Steln		Strelg		Strein	
Subject	m	sd	m	sd	m	sd	m	sd	m	sd	m	sd
A1	192	11.69	199	23.25	158	16.82	142	20.28	143	18.36	161	22.71
A2	249	29.77	234	19.56	208	26.82	160	15.57	179	25.69	219	34.63
A3	196	19.65	185	17.11	171	22.48	134	18.36	163	16.29	123	8.76
K1	207	14.76	227	14.98	236	41.28	212	31.25	241	25.15	163	26.48
	140	12.61	135	14.90	108	11.17	101	12.18	105	10.89	114	33.55
K2	250	34.47	222	13.89	198	27.75	208	24.42	198	38.06	177	20.52
	165	16.41	152	19.12	128	37.95	140	34.47	106	21.90	123	49.67
Hawkins (1976) KY1	108 - 191				140 - 181				134 - 219			
range of means KY2	178 - 211				151 - 194				141 - 224			

In perception data, sample size for adults = 11 or 12, and for children = 7 or 8 (see Table 2). In production data, sample size = 10 for A1, A2, A3, K1 and K2. It is ≥24<36 for each child in Hawkins (1976).

Table 6. Mean durations and standard deviations (ms) in perception and production for /t/ in a child's voice. Adult subjects: perception data only. Child subjects: upper row perception; lower row production. Hawkins (1976), production data only (KY1 7 child speakers; KY2 7 child speakers); see text for further explanation.

	Word and Context											
	Teig		Tein		sTeig		sTein		sTreig		sTrein	
Subject	m	sd	m	sd	m	sd	m	sd	m	sd	m	sd
A1	135	25.21	109	25.40	109	15.51	99	18.73	65	12.57	83	15.67
A2	109	9.93	71	6.16	120	20.21	47	6.51	83	13.34	73	10.98
A3	96	18.59	71	17.68	84	20.19	84	14.42	92	10.03	72	7.99
K1	-	-	-	-	112	18.60	106	3.71	104	11.35	97	7.33
	75	17.83	64	13.08	76	13.75	63	10.10	66	11.07	67	9.22
K2	-	-	-	-	95	15.96	100	5.66	105	11.86	96	10.08
	94	17.57	91	20.58	79	18.45	83	12.52	70	19.92	84	22.27
Hawkins (1976) range of means	KY1	35 - 97			50 - 77				51 - 85			
	KY2	50 - 96			67 - 102				62 - 116			

In perception data, sample size for adults = 11 or 12, and for children = 7 or 8 (see Table 2). In production data, sample size = 10 for A1, A2, A3, K1 and K2. It is 24-36 for each child in Hawkins (1976).

Table 7. Spearman rank correlation coefficients between perception and production durations

Subject	r_s	n	p*
A1	0.859	12	.01
A2	0.854	12	.01
A3	0.858	12	.01
K1	0.661	10	.05
K2	0.894	10	.01

* 1 tailed tests

Critical r_s at p 0.01 = 0.712 for n = 12 and 0.746 for n = 10.

Table 8. Kendall coefficients of concordance and Spearman rank correlations between subjects on durations in production and perception

Subjects	Production	Perception	Production and perception
Adults	0.984**	0.944**	0.908***
	26.57	25.49	49.03
Children	0.939**	0.694*	0.815***
			29.34
All subjects	0.942***	0.870***	0.861***
	42.39	39.15	77.49

Where there are two rows, the upper row W and the lower row = χ^2. Where there is one row ('children' on production or perception durations alone), the coefficient computed is r_s. For all coefficients, $n = 10$; d. f. for W and χ^2 is 9. Tests of significance are 1 tailed, because inverse relationships are not expected.

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 9. Analysis of variance for mean durations of responses expressed as deviations from the mean of the logarithms of each distribution (see text for explanation)

Source	Sum squares	df	Mean square	F
voices (V)	267.55	1	267.55	0.045
words (W)	1857.68	4	464.42	0.079
V x W	2406.59	4	601.65	0.102
carriers (C)	62819.94	1	62819.94	10.614***
V x C	231656.20	1	231656.20	39.140***
W x C	134963.10	4	33740.78	5.701***
V x W x C	197590.43	4	49397.61	8.346***
subjects (Sub)	1035.50	4	258.88	0.044
V x Sub	1369.07	4	342.27	0.058
W x Sub	7320.80	16	457.55	0.077
C x Sub	78788.04	4	19697.01	3.328***
V x W x Sub	14899.72	16	931.23	0.157
V x C x Sub	311178.80	4	77794.70	13.144***
W x C x Sub	638532.59	16	39908.29	6.743***
V x W x C x Sub	962840.51	16	60177.53	10.167***
Residual	3401399.90	923	5918.72	

*** significant at $p < 0.001$

Table 10. Marginals for voices x carriers (deviations from the mean logs)

		Voice	
		adult	child
Carrier	Now you'll hear _____ again	-6.856	+22.354
	Please do _____ now	+7.543	-23.444

Table 11. Marginals for voices x carriers x subjects (deviations from the mean logs)

	Adult voice		Child voice	
	carrier 1*	carrier 2**	carrier 1	carrier 2
A1	-3.848	+7.102	-9.617	+10.035
A2	-32.983	+35.000	+52.483	-51.150
A3	-4.067	+4.050	+33.466	-37.317
K1	+5.923	-4.615	+20.385	-25.500
K2	+9.950	-15.795	+10.000	-7.154

* 'Now you'll hear _____ again'

** 'Please do _____ now'

Table 12. Mann-Whitney U tests: Adult versus child listeners
Adult voice stimuli

Stimulus	n_1	n_2	U	z	p*	shortest durations set by
Selg	15	36	90.5	-3.71	.0002	children
Seln	15	36	161.5	-2.24	.025	"
Stelg	15	35	175.0	-1.85	.064	"
Steln	16	35			ns	
Strelg	16	36			ns	
Strein	16	36			ns	
sTelg	16	36			ns	
sTeln	16	36	71.0	-4.3	.00006	"
sTrelg	16	36			ns	
sTrein	15	36			ns	

Table 13. Mann-Whitney U tests: Adult versus child listeners
Child voice stimuli

Stimulus	n_1	n_2	U	z	p*	shortest durations set by
Selg	16	35	187.5	-1.88	.06	adults
Seln	16	36	161.0	-2.52	.01	"
Stelg	15	36	135.5	-2.78	.005	"
Steln	16	36	6.0	-5.59	.00006	"
Strelg	15	35	49.0	-4.52	.00006	"
Strein	16	35			ns	
sTelg	14	36			ns	
sTeln	16	36	107.5	-3.58	.0005	"
sTrelg	16	36	57.5	-4.57	.00006	"
sTrein	15	36	53.5	-4.47	.00006	"

* 2 tailed tests

Footnotes

1. Originally the stimulus set also included 'ray' and 'tray'. These were excluded after pilot experiments showed that subjects could not be expected to give the time required to complete the whole stimulus set. It was considered preferable to exclude stimuli rather than for a subject to perform on only a subset of stimuli. /r/ in 'tray' and 'stray' was excluded because as a liquid it is qualitatively different from /s,t/ (c.f. Lehiste (1972)). Word-initial /t/ was excluded in 'tray' and for children in 'tay' also, because subjects found it difficult and frustrating to deal with, as evidenced by their remarks and their large variances for this stimulus. This was because only the closure period was varied. It is very short; and therefore the cues to an adequate /t/ are mainly carried in the surrounding segments. Furthermore, subjects had a tendency to make it overlong, as if it were written with a colon before it, as "Now you'll hear: 'tay', again". An attempt was made to avoid this by manipulating the intonation contour, but this was only partly successful. Use of different carrier phrases could have reduced the problem still further, but since stimuli had to be dropped anyway, this was not done.

2. Nooteboom (1973) did report a difference of about 25ms. between means from initially long durations and those from initially short durations, but this is due to bias in the method of limits (see Criticism of the Method), and hence is not the same phenomenon as that under discussion here. In this experiment, where subjects were encouraged to move the wheel in both directions in a run of trials and to move past the first acceptable duration to 'check' their decision, there were no reliable differences dependent upon initial starting duration.

FIGURES

The following figures show spectrograms of the synthetic speech used in the experiments. The speech was created on an OVE III synthesiser by Sarah Hawkins using a PDP12 and SPEX as control. This work was carried out at University College London.

The spectrograms are standard Kay Sonagraph -type using the wide band (300Hz) filter and an approx. 8kHz frequency range.

Adult Male Voice:

1. Now you'll hear say again.
2. Now you'll hear tay again.
3. Now you'll hear stay again.
4. Now you'll hear stray again.
5. Now you'll hear [illegible]y again. (alternative intonation, not used in the experiment)
6. Now you'll hear say again. (intonation used in the experiment - as Fig.1)
7. Please do say now.
8. Please do tay now.
9. Please do stay now.
10. Please do stray now.

Child's Voice:

11. Now you'll hear say again.
12. Now you'll hear tay again.
13. Now you'll hear stay again.
14. Now you'll hear stray again.
15. Now you'll hear say again. (alternative intonation, not used in the experiment)
16. Now you'll hear say again. (intonation used in the experiment - as Fig.11)
17. Please do say now.
18. Please do tay now.
19. Please do stay now.
20. Please do stray now.

All these sentences were used in the experiment, with the exception of the alternative intonation with 'Now you'll hear ...'. This was to be the pattern used, but was rejected for reasons explained in the text. It is included here for the record, because it is probably preferable except for the problem that caused it to be rejected. It is paired with the F_0 contour actually used simply for auditory comparison.

Fig.1 Now you'll hear say again. Adult male voice

Fig.2 Now you'll hear tay again. Adult male voice

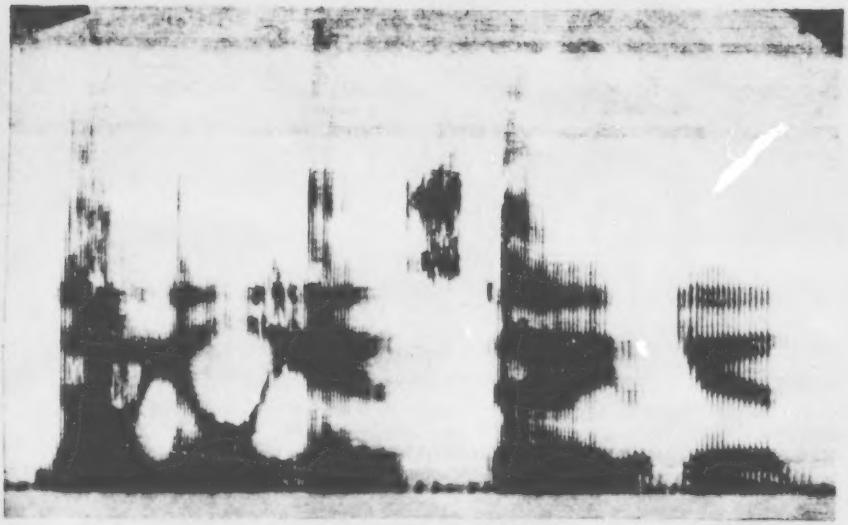

Fig.3 Now you'll hear stay again. Adult male voice

Fig.4 Now you'll hear stray again. Adult male voice

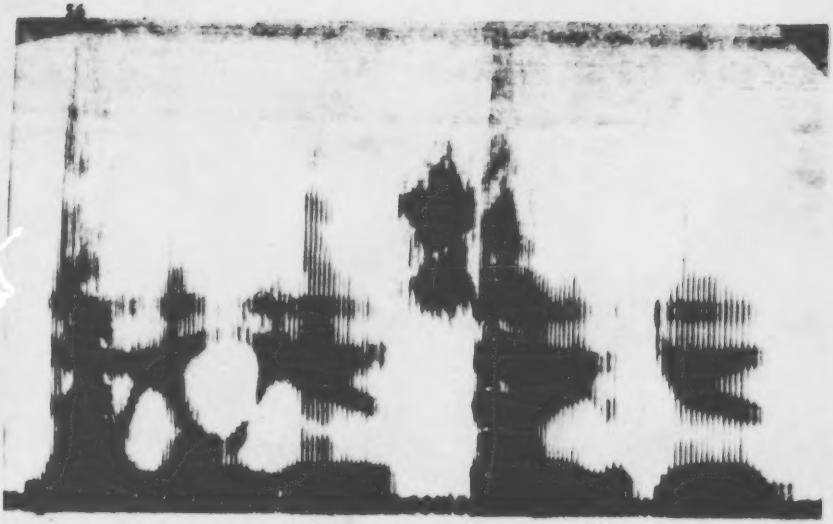

Fig.5 Now you'll hear say again. Adult male voice - not used

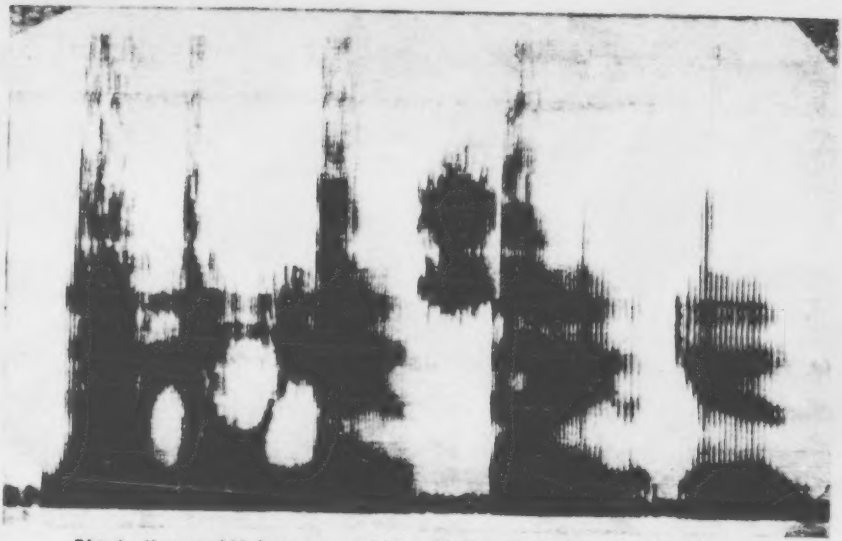

Fig.6 Now you'll hear say again. Adult male voice (as Fig.1)

Fig.7 Please do say now. Adult male voice

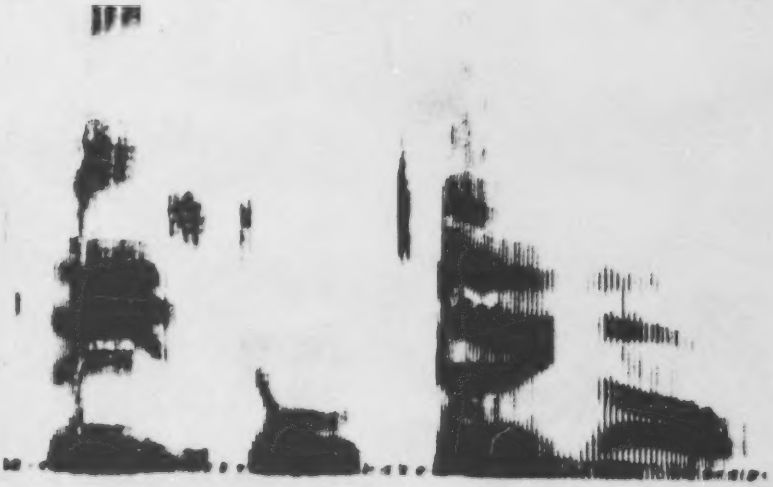

Fig.8 Please do tay now. Adult male voice

Fig.9 Please do stay now. Adult male voice

Fig.10 Please do stray now. Adult male voice

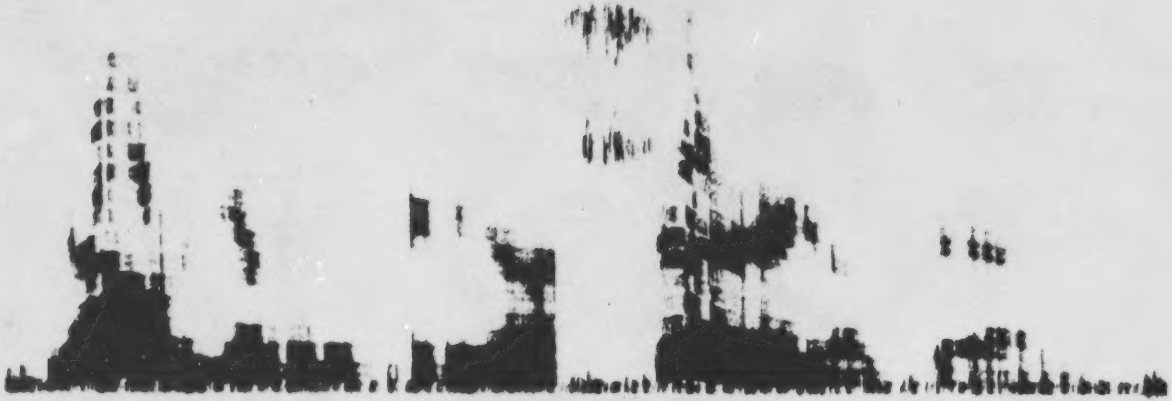

Fig.11 Now you'll hear say again. Child's voice

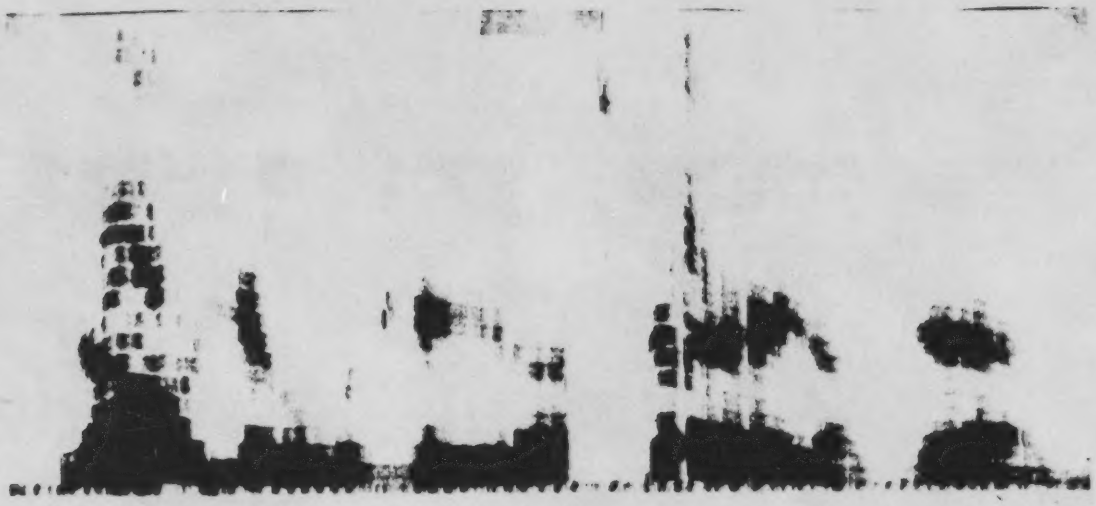

Fig.12 Now you'll hear tay again. Child's voice

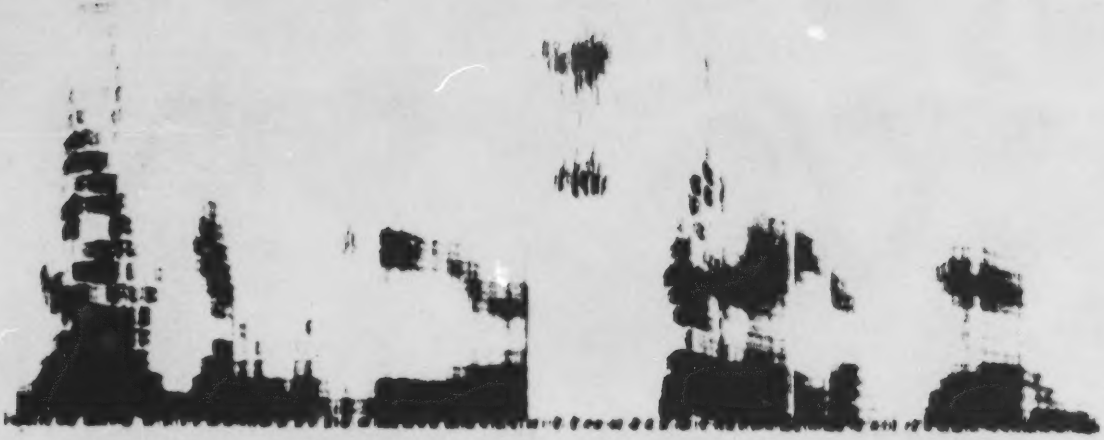

Fig.13 Now you'll hear stay again. Child's voice

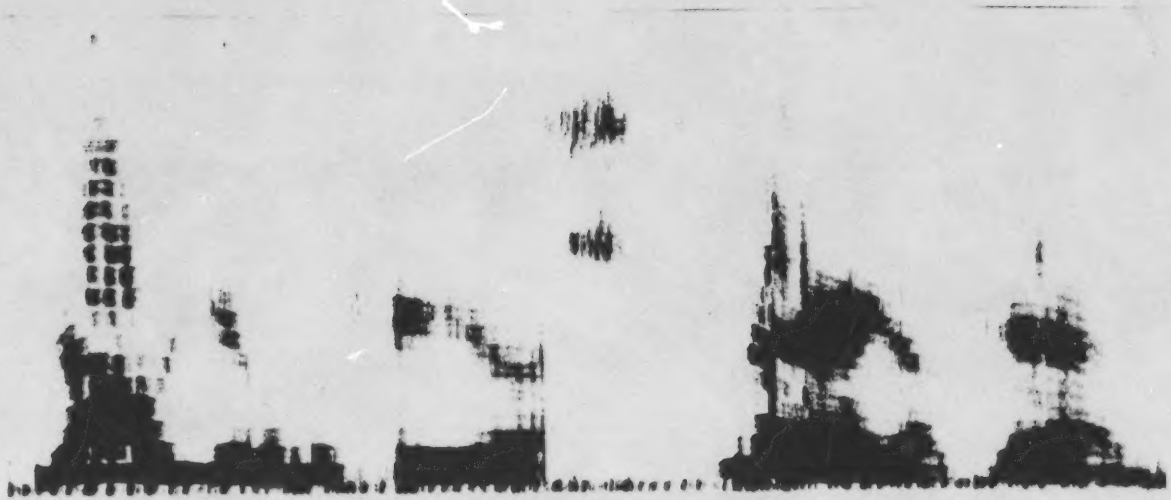

Fig.14 Now you'll hear stray again. Child's voice

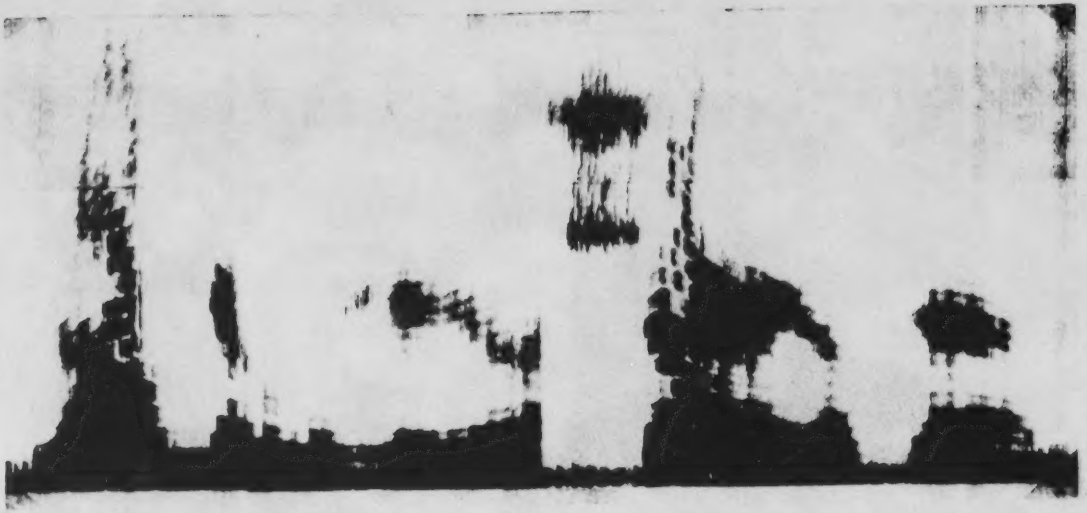

Fig.15 Now you'll hear say again. Child's voice - not used

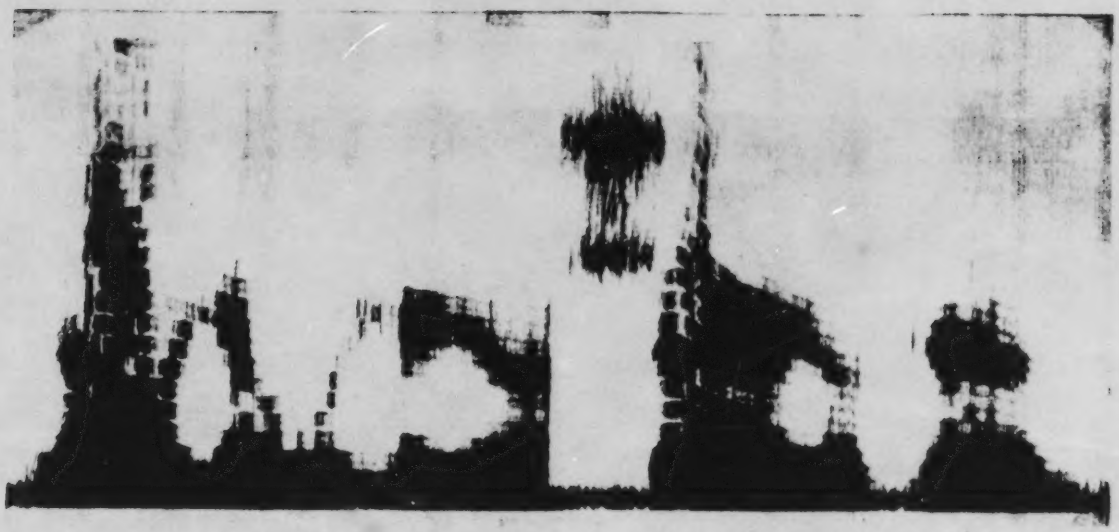

Fig.16 Now you'll hear say again. Child's voice (as Fig.11)

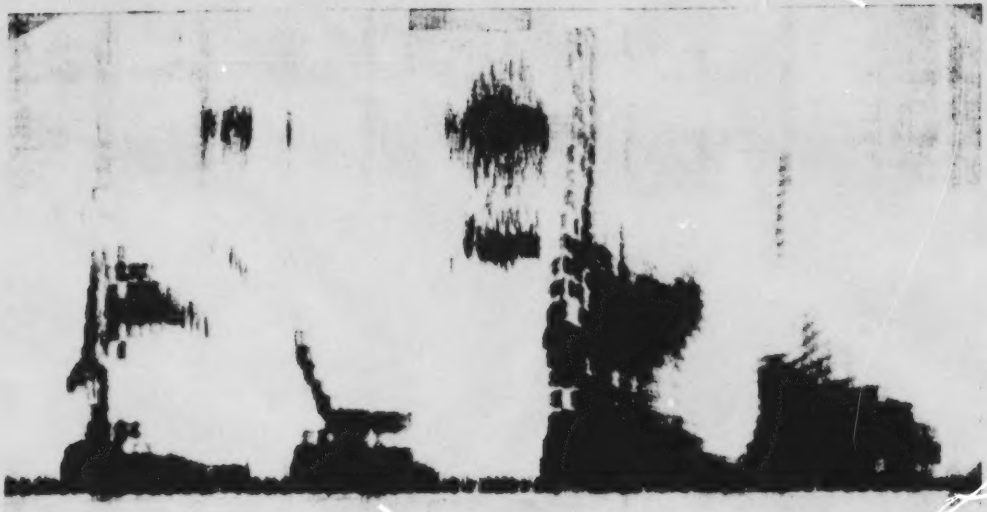

Fig.17 Please do say now. Child's voice

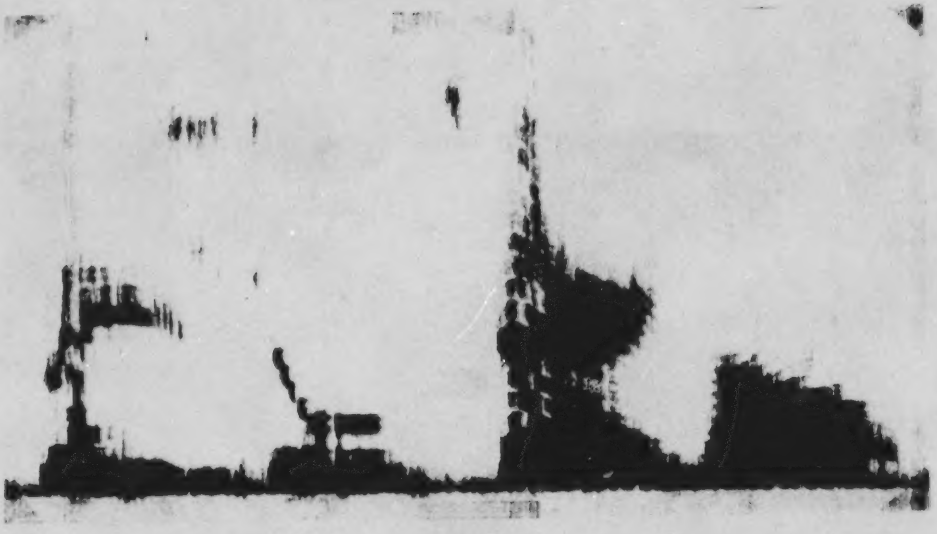

Fig.18 Please do tay now. Child's voice

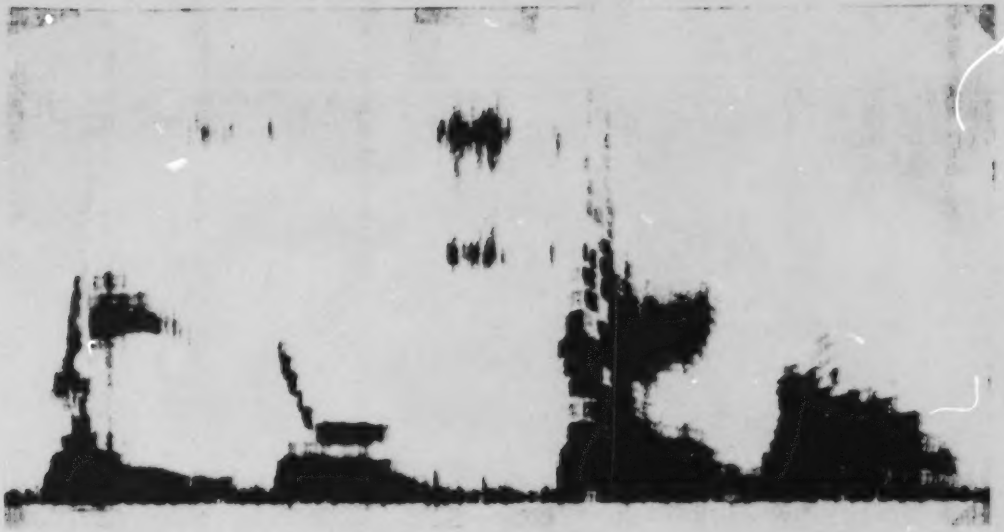

Fig.19 Please do stay now. Child's voice

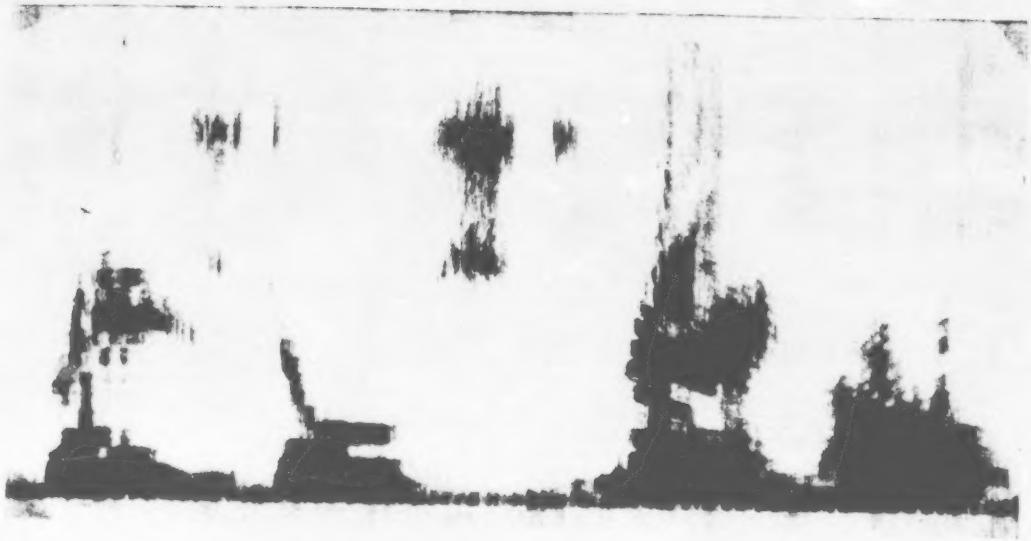

Fig.20 Please do stray now. Child's voice

Bibliography

Allen, G.D. (1972a) The location of rhythmic stress beats in English: an experimental study I. Language and Speech 15, 72 - 100.

Allen, G.D. (1972b) The location of rhythmic stress beats in English: an experimental study II. Language and Speech 15, 179 - 195.

Cartwright, D. (1941) Decision time in relation to the differentiation of the phenomenal field. Psychological Revue 48, 425 - 442.

Clark, H.H. (1973) The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior 12, 335 - 359.

Clark, H.H. (1976) Discussion of Wike and Church's comments: reply to Wike and Church. Journal of Verbal Learning and Verbal Behavior 15, 257 - 261.

Cohen, J. (1976) Discussion of Wike and Church's comments: random means random. Journal of Verbal Learning and Verbal Behavior 15, 261 - 262.

Creamer, L.R. (1963) Event uncertainty, psychological refractory period, and human data processing. Journal of Experimental Psychology 66, 187 - 194.

Davis, R. (1957) The human operator as a single channel information system. Quarterly Journal of Experimental Psychology 9, 119 - 129.

Draper, G. (1972) Speech perception experimental control system (SPECS). Speech Perception 2(1), 21 - 32. (Dept. of Psychology, The Queen's University of Belfast).

Draper, G. (1973) Speech perception experimental control system - progress report. Speech Perception 2(2), 39 - 43.

Fant, G. (1968) Analysis and synthesis of speech processes. In B. Malmberg (ed.) Manual of Phonetics. Amsterdam: North Holland Publishing Company.

Fant, G., Martony, J., Rengman, V. and Risberg, A. (1962) ØVE III synthesis strategy. Presented at the Speech Communication Seminar, Stockholm, Sweden, paper F5.

Forster, K.I. and Dickinson, R.G. (1976) More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F_1, F_2, F', and min F'. Journal of Verbal Learning and Verbal Behavior 15, 135 - 142.

Gladstone, V.S. and Siegenthaler, B.M. (1971) Carrier phrase and speech intelligibility test score. Journal of Auditory Research 11, 101 - 103.

Hawkins, S. (1976) The Developing Organisation of Speech Production in Children: evidence from consonant clusters. Unpublished PhD thesis, University of Cambridge.

Huggins, A.W.F. (1972a) Just noticeable differences for segment duration in natural speech. Journal of the Acoustical Society of America 51, 1270 - 1278.

Huggins, A.W.F. (1972b) On the perception of temporal phenomena in speech. Journal of the Acoustical Society of America 51, 1279 - 1290.

Keppel, G. (1976) Discussion of Wike and Church's comments: words as random variables. Journal of Verbal Learning and Verbal Behavior 15, 263 - 266.

Klatt, D.H. (1973a) Durational characteristics of prestressed word-initial consonant clusters in English. MIT Research Laboratory of Electronics: Quarterly Progress Report 108, 253 - 260.

Klatt, D.H. (1973b) Voice-onset time, frication and aspiration in word-initial consonant clusters. MIT Research Laboratory of Electronics: Quarterly Progress Report 109, 124 - 136.

Klatt, D.H. (1974) The duration of /s/ in English words. Journal of Speech and Hearing Research 17, 51 - 63.

Lehiste, I. (1970) Suprasegmentals. Cambridge, Mass: MIT Press.